Clinical Predictive Value of the *in Vitro* Cell Line, Human Xenograft, and Mouse Allograft Preclinical Cancer Models[1]

Theodora Voskoglou-Nomikos,[2] Joseph L. Pater, and Lesley Seymour

National Cancer Institute of Canada Clinical Trials Group, Cancer Clinical Trials Division, Cancer Research Institute, Queen's University, Kingston, Ontario, K7L 3N6 Canada

ABSTRACT

Purpose: We looked at the value of three preclinical cancer models, the *in vitro* human cell line, the human xenograft, and the murine allograft, to examine whether they are reliable in predicting clinical utility.

Experimental Design: Thirty-one cytotoxic cancer drugs were selected. Literature was searched for drug activity in Phase II trials, human xenograft, and mouse allografts in breast, non-small cell lung, ovary, and colon cancers. Data from the National Cancer Institute Human Tumor Cell Line Screen were used to calculate drug *in vitro* preclinical activity for each cancer type. Phase II activity *versus* preclinical activity scatter plot and correlation analysis was conducted for each model, by tumor type (disease-oriented approach), using one tumor type as a predictor of overall activity in the other three tumor types combined (compound-oriented approach) and for all four tumor types together.

Results: The *in vitro* cell line model was predictive for non-small cell lung cancer under the disease-oriented approach, for breast and ovarian cancers under the compound-oriented approach, and for all four tumor types together. The mouse allograft model was not predictive. The human xenograft model was not predictive for breast or colon cancers, but was predictive for non-small cell lung and ovarian cancers when panels of xenografts were used.

Conclusions: These results suggest that under the right framework and when panels are used, the *in vitro* cell line and human xenograft models may be useful in predicting the Phase II clinical trial performance of cancer drugs. Murine allograft models, as used in this analysis, appear of limited utility.

INTRODUCTION

Both basic science studies and clinical trials are essential components of the cancer drug discovery process. Potential therapeutics found to be significantly better than no treatment or standard therapies (*i.e.*, active) in preclinical laboratory cancer models or compounds with novel chemotypes and equivalent effectiveness to standard treatments are advanced to confirmatory testing in early (Phase I and II) clinical trials. Considering that RR[3] is a reasonable surrogate end point for survival (required but not sufficient), a favorable RR in Phase II trials advances a drug into additional clinical testing and is considered a prerequisite of drug success in the clinic.

Advancing of a candidate drug from preclinical testing in the laboratory to testing in Phase II clinical trials is based on the assumption that drug activity in cancer models translates into at least some efficacy in human patients, *i.e.*, that cancer laboratory models are clinically predictive. In addition, the relevance of tumor type-specific preclinical results for the corresponding human cancers in the clinic can be viewed through two different approaches: compound-oriented, where a drug is assumed to have potential activity against all human tumor types if it is effective against a single test tumor type, and disease-oriented, where a drug with preclinical activity in a single tumor type would only be expected to be effective in the same tumor type in patients.

Although widely adopted, the above-mentioned assumption and approaches have not been confirmed by studies to date. In addition, all studies aimed to examine the clinical predictive value of laboratory cancer models inevitably suffer from inherent bias because compounds with no activity in preclinical models are generally not advanced to clinical trials.

This work was undertaken to examine the clinical predictive value of three preclinical cancer models that have found wide use: the human *in vitro* cell line; the mouse allograft; and the human xenograft. In these models, tumor volume or life span (*in vivo* mouse models) or cell growth (*in vitro* cell lines) is compared between the treatment group receiving the new drug and a control group (active or inactive control).

The use of preclinical cancer models for selection of potential cancer therapeutics was pioneered by the NCI in the United States in the mid-1950s. The screening strategies used until 1990 were essentially compound oriented and involved a

---

---

[3] The abbreviations used are: RR, response rate; NCI, National Cancer Institute; NSCLC, non-small cell lung cancer; NSC, National Service Center; T/C%, treated over control tumor volume ratio.

small number of predominantly murine allograft tumors, with emphasis on leukemia (1–7). Several studies from the NCI and others demonstrated that this approach had low clinical predictive value for activity in Phase II trials (5–9) and yielded compounds with selective activity toward human leukemias and lymphomas (10–12). Thus, in 1990, the NCI introduced a disease-oriented *in vitro* Human Tumor Cell Line Screen comprised of 60 cell lines from the most common adult tumors (13–17). The screen was designed so that each tumor type was represented by a panel of cell lines, selected on the basis of different subhistological features, and common drug resistance profiles. It was hoped that this screen would help identify drug leads with high potency and/or selective activity against particular tumor types.

Recently, the NCI examined the correlation between drug activity in Phase II clinical trials and preclinical activity in cancer models (18). Important findings were: (*a*) with the exception of NSCLC, preclinical activity in human xenografts of a particular tumor type did not correlate significantly with Phase II activity in the same type of tumor, (*b*) with the exception of breast and colon histologies, human xenografts did not significantly predict Phase II clinical activity in other cancers types; and (*c*) compounds that were active in at least one-third of all tested human xenografts were likely to have at least some activity in Phase II clinical trials.

Studies examining the clinical predictive value of preclinical cancer models outside the scope of the NCI screening programs have focused on the human xenograft model and have looked predominately into same-tumor correlations (disease-oriented approach). These studies have produced both positive (the model was found clinically predictive) and negative (the model was found to have no clinical predictive value) results in various tumor types (19–27).

Two major criticisms can be made on the overall body of literature concerning the clinical predictive value of preclinical cancer models. First, the vast majority of studies to date, both within and outside the NCI, have based their conclusions on the observation of trends rather than the use of statistical methods. Second, all studies conducted previously have used dichotomous definitions of preclinical and/or clinical activity based on largely invalidated cutoff values of measures of activity: a 20% RR in Phase II clinical trials and (most commonly) a 42% T/C% in human xenografts and mouse allografts.

In addition, two important questions have not been addressed at all by previous studies: the clinical predictive value of the *in vitro* cell line model and the relative clinical usefulness of the different preclinical cancer models in use today (*i.e.,* how different models compare with each other in terms of their ability to identify clinically effective drugs).

Thus, we conducted a study comparing the clinical (Phase II) predictive value of three widely used preclinical laboratory cancer models, the *in vitro* human cell line, the mouse allograft, and the human xenograft. We used quantitative measures of both clinical and preclinical activity and statistical methods. We considered three relevant questions: (*a*) the clinical predictive value of the three models within the same tumor type (disease-oriented approach); (*b*) the clinical predictive value of the three models when one preclinical tumor type is used as a predictor of overall clinical activity in all other tumor types (compound-oriented approach); and (*c*) the clinical predictive value of the three models when overall preclinical and clinical activity in all tumor types combined is considered.

MATERIALS AND METHODS

Study Design

A retrospective, literature-based study was conducted. Data were retrieved from studies published between 1985 and 2000. This period was chosen as one when all three preclinical cancer models of interest to this study were in use and because it was long enough and close enough to the present as to afford data on a relatively large number of recently developed drugs.

The data search was restricted to four of the most common and commonly studied solid tumor types, breast, colorectal, ovarian, and non-small cell lung cancers, to ensure that sufficient data would be available.

The Medline and CancerLit databases were used for the collection of published data. In an attempt to minimize publication bias, both paper publications (peer reviewed) and meeting abstracts (nonpeer reviewed) were used as sources of information. If published data were not available for identified drugs, manufacturers were contacted for unpublished data.

Selection of Drugs

Drugs were identified by searching the Medline and CancerLit databases for compounds that had undergone single agent Phase I clinical trial testing either in 1991 or 1992. Agents with novel targets such as signal transduction or angiogenesis modulators were not included.

This Phase I-based approach to agent identification was used to ensure selection of agents developed within the study time frame of 1985–2000: agents with a published Phase I clinical trial in 1991 or 1992 were expected to have been through preclinical testing between 1985 and 1990 and to have undergone Phase II clinical evaluation by the year 2000. In addition, this approach was adopted to minimize publication bias: publication of Phase I trials is generally less dependent on the observation of favorable tumor responses than publication of Phase II trials or of preclinical cancer model experiments.

Data Collection and Drug Activity

Phase II Clinical Trials. Phase II clinical trials for each drug were identified by searching the Medline and CancerLit databases for scientific papers, reviews, or meeting abstracts. Duplicate publications were discarded. For trials with only abstract information, an additional search by author and/or institution name was conducted in Medline or CancerLit. Scientific papers were used in preference to abstracts, where possible.

Two restrictions were applied. The first was a geographic restriction: to ensure uniform methodology in trial conduct and RR assessment, only Phase II trials conducted in the Americas, Western Europe and Australia were included in the analysis. The second restriction referred to the treatment population and aimed to ensure that uniformly responsive populations of patients would be considered. For breast and ovarian cancer, only Phase II trials that included patients who had received prior chemotherapy for metastatic disease were used, whereas for

NSCLC and colon cancers, the Phase II trials selected included patients who had received no prior chemotherapy.

For each individual Phase II trial the following information was collected: disease site; previous chemotherapy; disease stage; number of patients entered; eligible; evaluable and evaluable for response; number of complete and partial responses; and criteria used for response (standard WHO *versus* other). Trials had to have enrolled a minimum of 14 patients, at least 12 of whom must have been evaluable for response. Completed Phase II trials for which >20% of entered patients were listed as inevaluable for response were considered methodologically unacceptable and were not used. For trials in progress at the time of reporting (meeting abstract format only), the available data were used even if they represented <80% of the enrolled patients, provided that they met the 14-patient criterion. If a trial publication did not specify the previous chemotherapy treatment status of patients, it was not used. Information from Phase I-II trials was used only when the Phase I and II components of the trial were separately conducted and reported. Phase II information was collected regardless of drug dose and route of administration.

For a given drug, in a given cancer type, the activity in a single Phase II clinical trial was recorded as the RR: the number of partial and complete tumor responses over the total number of patients evaluable for response. The number of evaluable rather than eligible patients was used to accommodate information from trials for which final results were not available. In the very few cases where the number of patients evaluable for response was not provided, the number of evaluable patients, the number of eligible patients, or the number of patients entered in the trial (whichever was provided by the investigators) in that priority order was used.

To obtain a drug's overall clinical activity in multiple Phase II trials of patients with the same tumor type, all responses and the collective number of patients evaluable for response were pooled from individual trials to calculate an overall RR. Finally, to get the Phase II activity for any three or four cancer types combined, the individual tumor RRs were averaged.

Human Xenografts and Mouse Allografts. The search strategy for mouse cancer model data were similar to the Phase II process. The only exclusion in this case were results obtained with mouse tumors that were engineered to have special characteristics such as, for example, overexpression of proteins conferring drug resistance.

For each murine allograft or human xenograft, numerical value(s) of activity for drugs of interest was retrieved only if expressed as the treated over control tumor volume ratio (T/C%) or the tumor volume growth inhibition ratio (GI%; and T/C% = 100% − GI%) in the literature sources. In addition, only T/C% values calculated by the formula T/C% = [(RVtreated)/(RV-control)] × 100% were collected (where RV = relative volume), whereas T/C% values defined for regressions [T/C% = [(RVtreated (0) − RVtreated(t))/RVtreated (0)] × 100%] were excluded to ensure uniform calculation methods. If the T/C% was not provided but a relative tumor growth curve was given as a figure in a publication, the numerical values for the treatment and control groups provided in this graph were used to calculate the T/C%. Activity reported as all mice cured or 100% complete responses was considered equivalent to and recorded as a T/C%

= 0. If no exact T/C% value was given but an interval of values was provided instead (*i.e.*, T/C% >42), a T/C% equal to the interval midpoint value (*i.e.*, a T/C% = 71) was assigned. Finally, where preclinical activity was reported as GI%, it was converted to T/C% by the formula T/C% = 100% − GI%. The activity value for the most effective, nontoxic dose in each schedule was recorded.

Single tumor type preclinical activity of each drug in the murine allograft or human xenograft models was defined as the mean T/C% value from all tested allografts/xenografts of that tumor type. Where the same laboratory had tested a single xenograft/allograft with multiple schedules of the same drug and/or where the same xenograft/allograft had been tested with the same drug by more than one laboratories, T/C% values for a single tumor were obtained by first averaging the same laboratory T/C% values and then the same xenograft T/C% values.

Overall preclinical activity in xenografts/allografts for all four tumor types together was expressed as the average of single tumor mean T/C% values.

In Vitro **Human Tumor Cell Lines.** The publicly available data from the NCI's Human Tumor Cell Line Screen was used as the information source for the *in vitro* tumor cell line model. Information from the NCI *in vitro* Human Tumor Cell Line Screen was favored because it was a readily available, well-defined, comprehensive, validated, and extensive single source of data. Another important reason was that as an exploratory literature search showed, there was such a wide variation between different investigators in the types of assays used and the nature of cell lines tested that it would have been impossible to comprehensively combine published data from various laboratories.

Acquisition of NCI Human Tumor Cell Line Screen data were done through the internet.[4] Information for each drug was obtained through its NCI code number or NSC number. Such numbers, where available, were identified either from the literature or from a cross-reference of compound names and NSC numbers in the NCI database (also available on the NCI web site).[4]

Testing of compounds in the NCI *in vitro* Human Tumor Cell Line Screen has been described previously (17). Briefly, growth inhibition in cell lines is measured by the GI_{50}, defined as the drug concentration that causes a 50% reduction in cell number in test plates relative to control plates. For every drug entering the screen, a concentration range comprised of five, 10-fold dilutions is tested in each of a group of 60–80 cell lines. The optical densities between treated and control plates, as resulting from the sulforhodamine B assay, are used to construct a dose-response curve for each cell line in the screen, leading to the calculation of a GI_{50} in every case by interpolation. In the case of compounds with low (*i.e.*, the highest concentration tested causes <50% growth inhibition) or high (*i.e.*, the lowest concentration tested causes >50% growth inhibition) potency where interpolation is not possible, the highest and lowest concentrations, respectively, in the tested drug concentration

---

range are recorded as the approximated $GI_{50}s$. $GI_{50}s$ are then converted to their Log_{10} values and the overall mean $Log_{10}GI_{50}$ across all cell lines in the screen is calculated. Finally, the results are displayed by a bar graph called the mean graph (28). This graph lists all of the cell lines and their corresponding $Log_{10}GI_{50}s$ and relates the magnitude of every individual cell line $Log_{10}GI_{50}$ to the mean $Log_{10}GI_{50}$ across all of the cell lines by a bar to the right (more sensitive than average) or to the left (less sensitive than average) of a vertical line. The experiment is repeated several times for each concentration range. In cases where mean graphs are based on mostly approximated $GI_{50}s$, other higher or lower concentration ranges of the drug (again made of five, 10-fold dilutions) are also tested. Thus, for each compound tested in the NCI *in vitro* Human Tumor Cell Line Screen, multiple GI_{50} mean graphs (one for each concentration range tested) based on multiple experiments each and with a different content of approximated *versus* calculated (by interpolation) $GI_{50}s$ may exist in the NCI database.

We obtained all of the available GI_{50} mean graph information from the NCI web site for all drugs in our list of compounds with known NSC numbers.[4] For every drug, we recorded the number of concentration ranges tested in the NCI *in vitro* Human Tumor Cell Line Screen, the number of experimental repetitions conducted for each concentration range, and, finally, the number of approximated $Log_{10}GI_{50}s$ in each mean graph.

The drug concentration range that produced the mean graph with the smallest number of approximated $Log_{10}GI_{50}s$ was used for scoring a drug's activity in the NCI *in vitro* Human Tumor Cell Line Screen, unless a different concentration range existed, with a number of approximated $Log_{10}GI_{50}s$ varying <10% from the first but for which more experiments were done.

Preclinical activity in the NCI *in vitro* Human Tumor Cell Line Screen was scored in two different ways: by the mean $Log_{10}GI_{50}$ and by what was termed the activity fraction. For a given drug, in a given tumor type, the mean $Log_{10}GI_{50}$ was computed by averaging the $Log_{10}GI_{50}s$ from all of the cell lines of that tumor type in the mean graph corresponding to the most appropriate concentration range. The activity fraction was arbitrarily defined as the number of cell lines of a given tumor type in which the individual $Log_{10}GI_{50}s$ were more sensitive to the drug than the average $Log_{10}GI_{50}$ (for all cell lines of all cell types) in the mean graph over the total number of cell lines tested from that tumor type. The activity fraction was also calculated from the mean graph corresponding to the most appropriate concentration range. Overall mean $Log_{10}GI_{50}s$ or activity fractions for all four cancer types combined were calculated by averaging the single tumor values.

Statistical Analysis

For each preclinical cancer model, 9 Phase II *versus* preclinical activity relationships were examined for a total of 27: relationships by tumor type (disease-oriented approach, 4 relationships/model), predictive ability of one tumor type for the other three tumor types combined (compound-oriented approach, 4 relationships/model), and general predictive ability for all four tumor types combined (1 relationship/model).

Relationships were first examined descriptively with the construction of various Phase II overall activity *versus* preclin-

Table 1 Drugs selected for data collection. NSC numbers are shown, where available

Drug	NSC number
Taxotere	628503
Paclitaxel	125973
Topotecan	609699
Irinotecan	
Rhizoxin	332598
Gemcitabine	
Fazarabine	281272
Teniposide	122819
Menogaril	269148
Fosquidone	D611615
Elsamitrucin	369327
Amonafide	308847
Didemnin B	325319
Suramin	
Raltitrexed	639186
Flavone acetic acid	347512
Epirubicin	256942
CI-921	343499
Trimetrexate	352122
Multitargeted antifol	
Vinorelbine	
Piritrexim	351521
Fotemustine	
CI-980	
Chloroquinoxaline sulfonamide	339004
Ilmofosine	
CI-941	
Tiazofurin	286193
Pyrazine diazohydroxide	361456
Tallimustine	
Crisnatol	

ical activity scatter plots (Microsoft Excel software). Each point on these scatter plots represented data from one drug for which both Phase II and preclinical activity values had been calculated from literature sources, as described above.

After descriptive evaluation of the data, Spearman rank correlation coefficients were obtained using the SAS software, UNIX version 6.12. A significance test of every correlation coefficient was performed, and the corresponding *P*s were calculated. Spearman rank (nonparametric) correlation coefficients were used because the distributions of the *x* (preclinical activity) and *y* (clinical activity) variables were not normal (29).

When multiple comparisons are made within a group of data such as in this work, there is increased possibility that some correlations will come up as statistically significant solely because of chance (false positives). To avoid this, multiple comparison correction methods (*e.g.*, Bonferroni approach) are often used to adjust the significance level to a lower *P* than conventionally used. However, relying on corrected probabilities increases the possibility that meaningful correlations will be missed (false negatives), making the nature of the scientific work key to the decision to use multiple comparison adjustment methods or not. Because this was an exploratory study, we were willing to accept a higher probability of false positives to ensure that potentially meaningful associations would not be discarded. We therefore did not correct for multiple comparisons and chose a level of significance of 0.05.
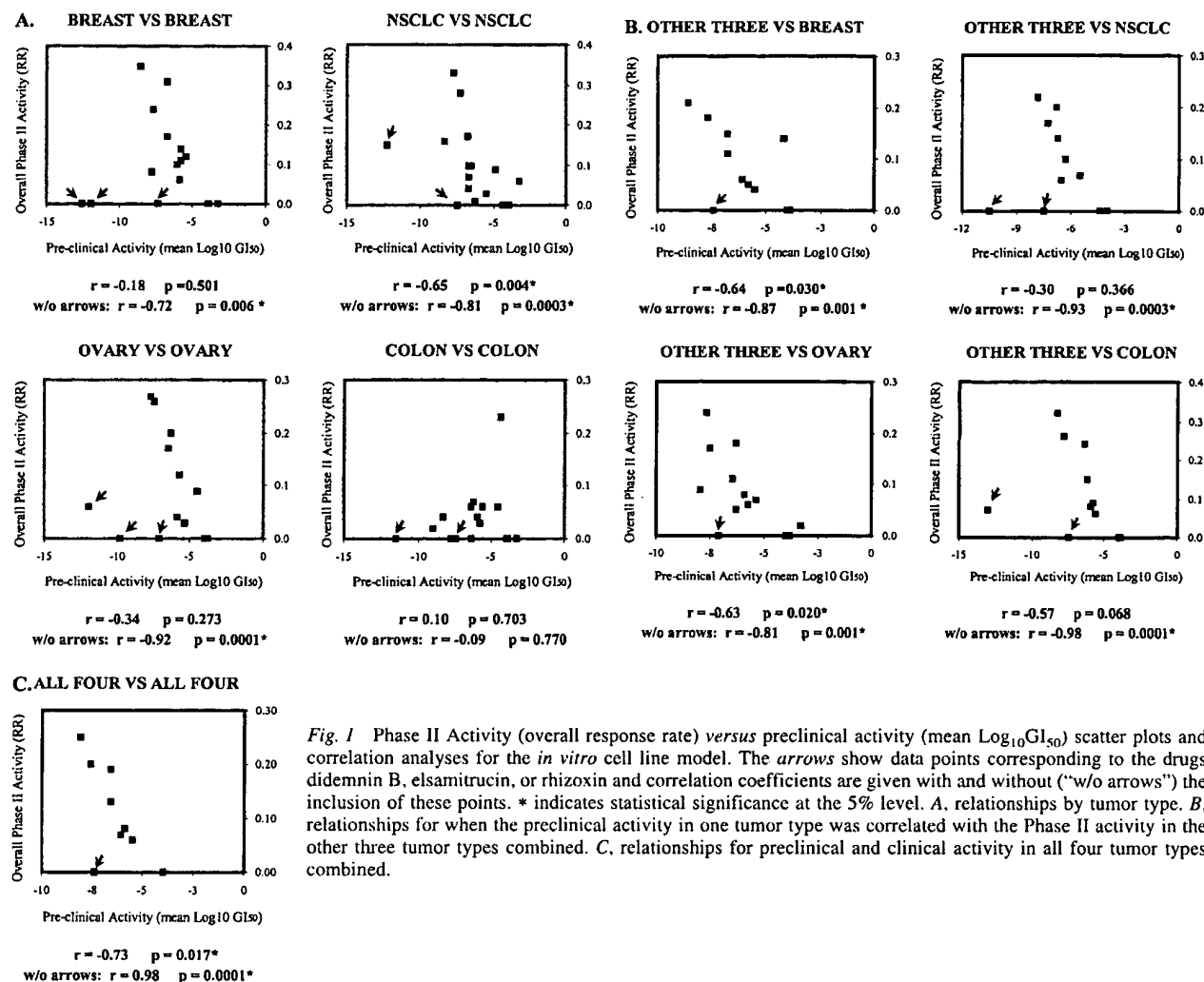
Fig. 1 Phase II Activity (overall response rate) *versus* preclinical activity (mean $Log_{10}GI_{50}$) scatter plots and correlation analyses for the *in vitro* cell line model. The *arrows* show data points corresponding to the drugs didemnin B, elsamitrucin, or rhizoxin and correlation coefficients are given with and without ("w/o arrows") the inclusion of these points. * indicates statistical significance at the 5% level. *A*, relationships by tumor type. *B*, relationships for when the preclinical activity in one tumor type was correlated with the Phase II activity in the other three tumor types combined. *C*, relationships for preclinical and clinical activity in all four tumor types combined.

RESULTS

The Medline and CancerLit databases were searched for cancer drugs (excluding agents with novel targets such as signal transduction or angiogenesis modulators) that had undergone single agent Phase I clinical trial testing either in 1991 or 1992. This search led to 97 drug names. After excluding drugs that were eliminated from additional clinical testing for practical reasons (for example difficulties with the drug formulation), drugs that were specifically developed for a certain type of cancer (as for example hormone-regulating compounds for breast cancer) and drugs that were still the subject of published Phase I studies in 1991 and 1992 despite already being licensed for human use before 1985, a list of 31 agents was obtained (Table 1). After applying the restrictions and criteria mentioned under "Materials and Methods," we extracted from the literature preclinical and Phase II activity information for those agents on four common cancer types, breast, NSCLC, ovary, and colon. Overall, 100 preclinical and 307 Phase II clinical literature references were used spanning the period between 1985 and 2000.

No preclinical data were found for 5 of the 31 drugs researched. Of the 26 drugs remaining, availability of preclinical and Phase II data varied, depending on which preclinical and clinical tumor(s) had been tested and published in each case. Thus, each of the relationships examined had a different number of data points as different subsets of drugs were included. The most data points for any relationship were 17. For six relationships, five or fewer data points were available (relationships with fewer than five data points were not included in the results presented below).

In Vitro **Cell Line Model.** Fig. 1 shows the Phase II activity *versus* preclinical activity scatter plots and correlation analysis for the *in vitro* cell line model when the mean $Log_{10}GI_{50}$ was used as the measure of preclinical activity. Because the lower the mean $Log_{10}GI_{50}$, the higher the potency of a drug, a negative correlation between mean $Log_{10}GI_{50}$ and Phase II overall RR was expected if the model had a good clinical predictive value. Significant negative correlations were found for NSCLC (Fig. 1*A*), for breast or ovarian cell lines *versus* overall Phase II activity in the other three tumor types
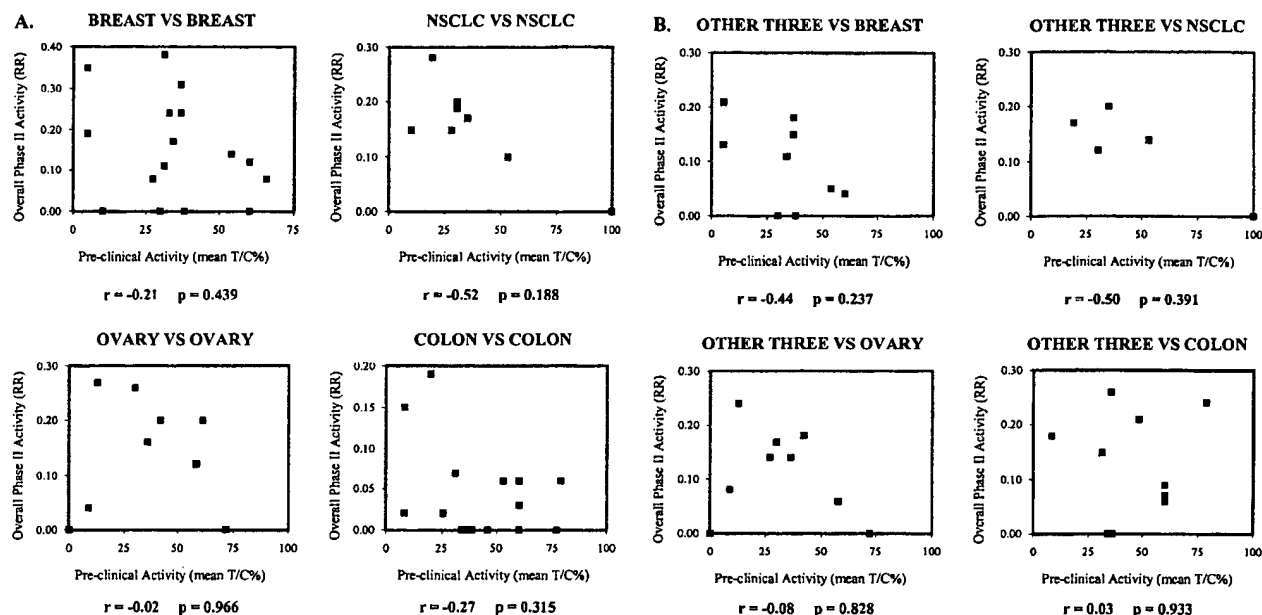
Fig. 2 Phase II activity (overall response rate) *versus* preclinical activity (mean T/C%) scatter plots and correlation analyses for the human xenograft model. *A*, relationships by tumor type. *B*, relationships for when the preclinical activity in one tumor type was correlated with the Phase II activity in the other three tumor types combined.

(Fig. 1*B*), and for preclinical activity *versus* Phase II activity in all four tumor types (Fig. 1*C*).

Although the trends observed with the activity fraction were similar to ones seen for the mean $Log_{10}GI_{50}$ measure, no correlations were statistically significant in this case (data not shown).

Human Xenograft Model. A negative correlation between Phase II RRs and mean T/C% values was expected to be indicative of a good clinical predictive value for the human xenograft model. As shown in Fig. 2, no significant correlations between preclinical and clinical activity were observed for this model in our analysis.

For some of the drugs, preclinical activity calculations were based on multiple human xenografts of the same tumor type (*i.e.*, panels) while for others on only a single xenograft. The relationships in Fig. 2 were reanalyzed, including only the drugs for which preclinical information on more than one human xenograft was available (Fig. 3). The results did not change for breast or colon tumors (compare Fig. 3*A* with Fig. 2*A*). However, the relationship for NSCLC became statistically significant and a highly significant correlation was seen for ovarian cancer (Fig. 3*A*). A near significant correlation was obtained when ovarian human xenograft panels were used to predict clinical activity in the other three tumor types combined (Fig. 3*B*).

Murine Allografts. No significant correlations between preclinical and clinical activity were observed for any of the relationships examined in this study for the murine allograft model (data not shown).

Additional Analyses. The scatter plots in Fig. 1 revealed an interesting observation: in every relationship except for colon

cancer under the disease oriented approach, an obvious trend toward a negative correlation was evident except for one to three outlier data points (Fig. 1, *arrows*). Interestingly, in all cases, these outlier data points corresponded to the same three drugs, namely elsamitrucin, didemnin B, and rhizoxin.

In an attempt to provide a possible explanation for this observation, we considered the mechanism of action of all drugs that were included in the correlations in Fig. 1. From a total of 18 drugs (Table 2), 5, namely, elsamitrucin, didemnin B, rhizoxin, flavone acetic acid, and fosquidone, were distinct in that they seemed to act through mostly unknown pathways that were not the typical DNA-based mechanisms of action of cytotoxic cancer agents. Thus, although flavone acetic acid and fosquidone fitted the rest of the data, there seemed to be a plausible mechanistic basis for the outlier behavior of the data points for elsamitrucin, didemnin B, and rhizoxin. In fact, exclusion of these three drugs led to highly significant correlations in all cases except for the same tumor relationship in colon cancer (Fig. 1, correlation coefficients and *P*s for "w/o arrows"). It should be noted that none of the relationships examined for the human xenograft models (Figs. 2 and 3) included elsamitrucin, didemnin B, or rhizoxin as data points.

Because of the intriguing results obtained with the human NSCLC and ovarian xenograft panels in Fig. 3*A*, a more detailed examination of these panels was pertained. As seen in Figs. 4*A* and 5*A*, the 6 ovarian and 7 NSCLC xenograft panels differed both in the numbers (minimum of 6 and maximum of 13 for ovary and minimum of 2 and maximum of 8 for NSCLC) and the identity of the xenografts that they contained. Analysis by grade/histology was hindered by lack of complete information on all xenografts. However, some patterns appeared distinguish-
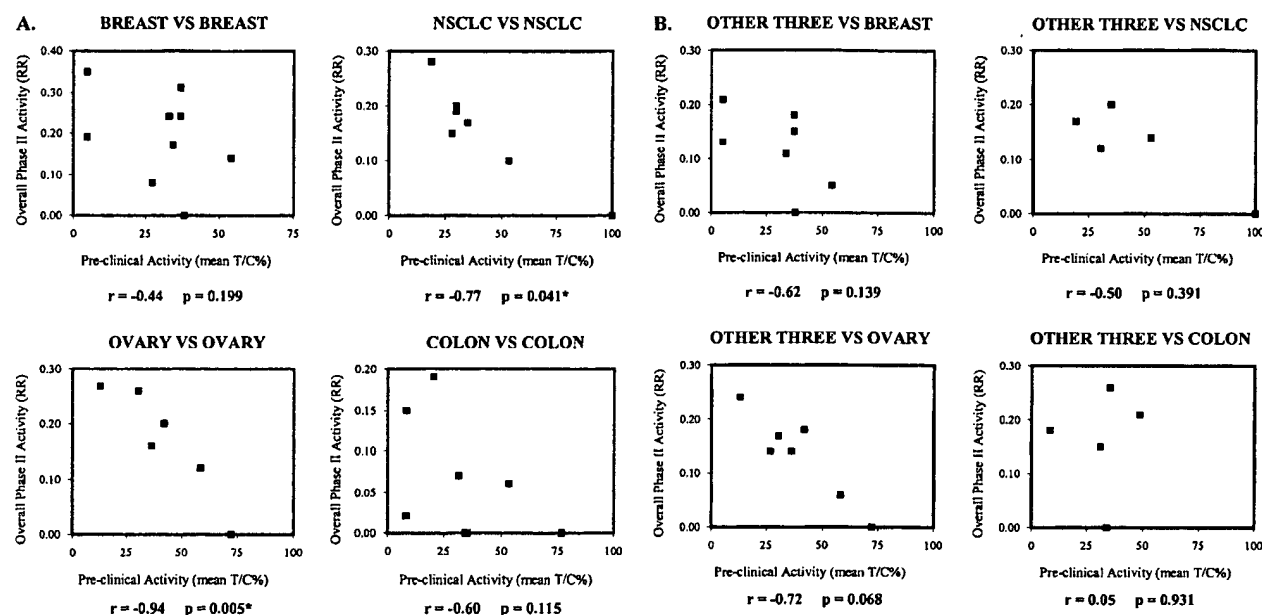
A. BREAST VS BREAST NSCLC VS NSCLC **B.** OTHER THREE VS BREAST OTHER THREE VS NSCLC

OVARY VS OVARY COLON VS COLON OTHER THREE VS OVARY OTHER THREE VS COLON

Fig. 3 Phase II activity (overall response rate) *versus* preclinical activity (mean T/C%) scatter plots and correlation analyses for the human xenograft model. Only data points for which two or more human xenografts were used to generate the preclinical activity values are shown. * indicates statistical significance at the 5% level. *A*, relationships by tumor type. *B*, relationships for when the preclinical activity in one tumor type was correlated with the Phase II activity in the other three tumor types combined.

Table 2 Mechanisms of action of drugs used in clinical vs. pre-clinical correlations for the *in vitro* cell line model (Fig. 1) Atypical cytotoxics are shown in bold.

Drug	Mechanism of action
Amonafide	DNA intercalator
CI-921	Acts on topoisomerase II
Didemnin B	Not understood. Believed to act on protein synthesis
Elsamitrucin	Not understood. It has been observed to inhibit topoisomerase I and II in *in vitro* experiments (relevance to *in vivo* uncertain). In cells in culture it has been observed to cause a cytostatic effect.
Epirubicin	Attaches to DNA at G bases
Fazarabine	Probably inhibits DNA synthesis by incorporation into DNA.
Flavone acetic acid	Has antivascular action in mice (probably not applicable to humans). Also believed to induce cell cycle arrest by generating reactive oxygen species that act on DNA.
Menogaril	Causes cleavage of double-stranded DNA by inhibiting topoisomerase II
Piritrexim	Inhibits dihydrofolate reductase
Rhizoxin	Not fully understood. May interact with tubulin (different binding site than taxoids) and lead to cell cycle arrest. Also observed to act as an angiogenesis inhibitor.
Taxol	Microtubule destabilizing agent that causes apoptosis
Taxotere	Microtubule destabilizing agent that causes apoptosis
Teniposide	DNA synthesis inhibition by stabilization of cleavable DNA complexes
Topotecan	Topoisomerase I inhibitor
Trimetrexate	Antifolate
Fosquidone	Unknown. Not a DNA binder or a topoisomerase inhibitor
Tomudex	Thymidylate synthase inhibitor
Tiazofurin	Inhibits 5'-phosphodehydrogenase, the rate-limiting enzyme for guanine ribonucleotide synthesis

able. All ovarian panels contained 10–20% undifferentiated tumors and also included both poorly differentiated and moderately differentiated subtypes (Fig. 4*B*). For NSCLC, all panels included adenocarcinoma xenografts with a frequency of >30% (Fig. 5*B*). These observations suggested that the frequency of histological/grade subtypes within a xenograft panel may be an

important determinant of clinical predictivity rather than the number or the nature of the xenografts.

In an attempt to explore this hypothesis and to further examine the validity of the results obtained for ovarian cancer and NSCLC in Fig. 3*A*, the literature was reviewed for additional data. Six more agents with known overall Phase II RRs in

A.

NAME	HISTOLOGY / GRADE	DATA POINTS (DRUGS)					
		EPIRUBICIN	FOSQUIDONE	GEMCITABINE	MENOGARIL	TAXOTERE	PACLITAXEL
MRI-H-207	undifferentiated	+	+		+	+	
A2780	undifferentiated	+		+	+		+
Ov.He	mod. diff., mucinous	+	+	+	+		
Ov.Me	carcinosarcoma	+	+		+		
OvRiC	mod. diff., serous	+	+	+	+		
Fma	poorly diff., mucinous	+	+		+	+	
Ov.Pe	mod. diff., mucinous	+	+	+	+	+	
Fco	clear cell sarcoma	+	+		+		
T17	cystoadenocarcinoma	+					
T385	adenocarcinoma	+					
OvGR	mod. diff., mucinous		+				
Fko	mod. diff., serous		+	+		+	
OvG1	poorly diff., serous		+				
OVCAR-3	adenocarcinoma			+		+	+
A121a	?					+	+
HOC18	poorly diff., serous					+	+
HOC22	poorly diff., serous					+	+
A2780/DDP	undifferentiated						+
A2780/DX	undifferentiated						+
SKOV-3	adenocarcinoma						+
1° ovary 1	cystoadenocarcinoma						+
1° ovary 2	dediff. serous adenoc.						+
IGROV 1	moderately diff.						+
OVCAR-8	poorly diff. adenoc.						+
OVCAR-5	adenocarcinoma						+
OvSh	poorly diff., serous					+	
HOC22-S	poorly diff., serous					+	
TOTAL NO.		10	10	6	8	10	13

Fig. 4 Human ovarian xenograft panels for the six data points (drugs) used in the "Ovary *versus* Ovary" relationship in Fig. 3*A. A*, names and histology/grade (? = unknown, mod. diff. = moderately differentiated, poorly diff. = poorly differentiated, dediff. = dedifferentiated, adncrc = adenocarcinoma) of all of the xenografts tested. Inclusion of a particular xenograft in one of the panels is shown by a "+" sign in the corresponding row and under the appropriate drug column. *B*, histology/grade subtypes in the human ovarian xenograft panels by number and percentage.

B.

HISTOLOGY/GRADE FREQUENCIES IN HUMAN OVARIAN XENOGRAFT PANELS						
HISTOLOGY / GRADE	EPIRUBICIN NO. (%)	FOSQUIDONE NO. (%)	GEMCITABINE NO. (%)	MENOGARIL NO. (%)	TAXOTERE NO. (%)	PACLITAXEL NO. (%)
undifferentiated	2 (20)	1 (10)	1 (17)	2 (25)	1 (10)	3 (23)
mod. diff., mucinous	2 (20)	3 (30)	2 (33)	2 (25)	1 (10)	0 (0)
mod. diff., serous	1 (10)	2 (20)	2 (33)	1 (12.5)	1 (10)	0 (0)
poorly diff., mucinous	1 (10)	1 (10)	0 (0)	1 (12.5)	1 (10)	0 (0)
poorly diff., serous	0 (0)	1 (10)	0 (0)	0 (0)	4 (40)	2 (15)
unspecified	4 (40)	2 (20)	1 (17)	2 (25)	2 (20)	8 (62)
TOTAL	10 (100)	10 (100)	6 (100)	8 (100)	10 (100)	13 (100)

previously treated patients with ovarian cancer were found. Five and one of these compounds had been tested in a panel of 15 and 6 human ovarian xenografts, respectively (26, 30), which fitted the histology/grade patterns identified in Fig. 4*B*. Fig. 6*A* lists the names and Phase II RRs (31–56) of these additional drugs together with the six compounds that were included in the analysis in Fig. 3*A*. Fig. 6, *A* and *B*, also shows mean T/C% values scatter plots and statistical analyses for two cases: first, for when all of the available xenograft information was used, and second, for when mean T/C% calculations were based, where possible, on the arithmetically smallest panel, namely the one used for gemcitabine in Fig. 4. Highly significant correlations were obtained in both cases (Fig. 6*B*).

For NSCLC information on two additional agents was found: amsacrine [mean T/C% of 62 (26) and Phase II RR equal to 0.06 (31)] and doxorubicin [mean T/C% of 47 (26) and Phase II RR equal to 0.12 (32)]. Both had been tested in NSCLC human xenograft panels that included all three histological subtypes and had adenocarcinoma contents of 29 and 33%,

respectively. As for ovarian cancer, those two additional data points (Fig. 5*C, arrows*) enhanced the statistical significance of the relationship observed in Fig. 3*A*.

DISCUSSION

A literature-based, retrospective study was conducted to examine the clinical predictive value of three widely used preclinical cancer models, namely, the *in vitro* human tumor cell line, the human xenograft, and the murine allograft models. Four solid tumor types were selected, breast, NSCLC, ovary and colon, and data on a set of 31 anticancer agents (excluding agents with novel targets such as signal transduction or angiogenesis modulators) were collected. Preclinical activity in each model was correlated with RRs in Phase II clinical trials by tumor type (disease-oriented approach) in the case when one preclinical tumor type was used as a predictor of overall clinical activity in the other three tumor types combined (compound-oriented approach) and for all four tumor types together.

A.

XEN. NAME	XENOGRAFT HISTOLOGY	DRUGS						
		EPI	FAZ	GEM	IRINO	PACLIT	TOPO	VINRLB
T222	squamous cell	+						
T291	adenocarcinoma	+						
UCLA-P3	adenocarcinoma		+					
ACCOLU-78	squamouse cell		+					
NCI-H460	large cell			+	+	+	+	+
A549	adenocarcinoma			+	+	+	+	+
CaLu-6	adenocarcinoma			+				
H-74	?			+				
LC-376	?			+				
QG-56	squamous cell				+			+
NCI-H23	adenocarcinoma				+	+		
NCI-H226	squamous cell				+	+	+	
MV-522	adenocarcinoma					+		
CaLu-3	adenocarcinoma					+		
1° NSCLC	adenocarcinoma					+		
L2987	adenocarcinoma					+		
L-27	adenocarcinoma							+
LC-06	large cell							+
LU-65	large cell							+
PC-12	adenocarcinoma							+
LU-99	large cell							+
TOTAL NO.		2	2	5	5	8	3	8

Fig. 5 Human NSCLC xenograft panels for the seven data points (drugs) used in the NSCLC *versus* NSCLC relationship in Fig. 3*A*. *A,* drug names (EPI = epirubicin, FAZ = fazarabine. GEM = gemcitabine, IRINO = irinotecan, PACLIT = paclitaxel, TOPO = topotecan, VINRLB = vinorelbine) and histological subtypes (? = unknown) of all of the xenografts tested. Inclusion of a particular xenograft in one of the panels is shown by a "+" sign in the corresponding row and under the appropriate drug column. *B,* histological subtypes in the human NSCLC xenograft panels by number and percentage. *C,* scatter plot and correlation analysis for the same tumor clinical *versus* preclinical activity relationship in NSCLC, including the seven drugs in Fig. 6*A* as well as two additional agents, doxorubicin and amsacrine (data points shown with *arrows*), with known NSCLC Phase II and human xenograft activities.
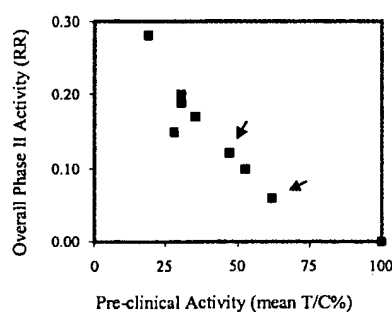
B.

HISTOLOGY FREQUENCY IN HUMAN NSCLC XENOGRAFT PANELS							
HISTOLOGY	EPI NO. (%)	FAZ NO. (%)	GEM NO. (%)	IRINO NO. (%)	PACLIT NO. (%)	TOPO NO. (%)	VINORLB NO. (%)
adenocarcinoma	1 (50)	1 (50)	2 (40)	2 (40)	6 (75)	1 (33.3)	3 (37.5)
large cell	0 (0)	0 (0)	1 (20)	1 (20)	1 (12.5)	1 (33.3)	4 (50)
squamous cell	1 (50)	1 (50)	0 (0)	2 (40)	1 (12.5)	1 (33.3)	1 (12.5)
unknown				2 (40)			
TOTAL	2 (100)	2 (100)	5 (100)	5 (100)	8 (100)	3 (100)	8 (100)

C. NSCLC VS NSCLC (ADDITIONAL DATA)

Pre-clinical Activity (mean T/C%)

r = -0.90   p = 0.001*

Colon cancer was the only site for which a disproportional amount of clinically active *versus* inactive agents were identified: only 3 drugs with Phase II RRs > 0.15 and 8 with ≤0.10 (Figs. 1–3). However, this was likely a reflection of the lack of clinically effective drugs for this tumor type rather than the result of selection and publication bias.

When the mean $Log_{10}GI_{50}$ measure of preclinical activity was used, the *in vitro* cell line model was found to be predictive of Phase II clinical performance for NSCLC under the disease-oriented approach in breast and ovarian cancers under the compound-oriented approach and in the case of all four tumor types together. Highly significant correlations were observed in all cases, except colon cancer, when three consistent outlier data points corresponding to the mechanistically nontypical cytotoxic agents didemnin B, elsamitrucin, and rhizoxin were excluded in exploratory analysis. Thus, the *in vitro* cell line model

A.

DRUG		PHASE II RESPONSE RATE	HUMAN OVARIAN XENOGRAFT MEAN T/C%	
			ALL TESTED	GEMCITABINE PANEL
STUDY DRUGS	EPIRUBICIN	0.20	42	-
	FOSQUIDONE	0.00	72	-
	GEMCITABINE	0.16	36	36
	MENOGARIL	0.12	58	-
	PACLITAXEL	0.26	30	-
	TAXOTERE	0.27	13	-
ADDITIONAL DRUGS	DOXORUBICIN	0.19 [13,16]	47 [20]	47 [20]
	AMSACRINE	0.05 [31]	75 [26]	-
	CISPLATIN	0.25 [37,38]	41 [30]	46 [30]
	HEXAMETHYL-MELAMINE	0.19 [39,44]	28 [30]	31 [30]
	METHOTREXATE	0.09 [45,46]	76 [30]	84 [30]
	5-FU	0.10 [47,46]	71 [30]	71 [30]

B.

OVARY VS OVARY (ALL)

r= -0.86 p= 0.0004*

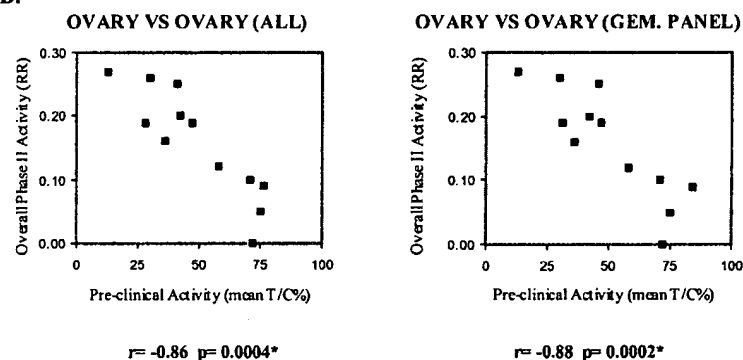OVARY VS OVARY (GEM. PANEL)

r= -0.88 p= 0.0002*

Fig. 6 *A*, preclinical and Phase II clinical activity data for ovarian cancer, including the six drugs in Fig. 3A ("Study Drugs") as well as an additional six drugs ("Additional Drugs") with known ovarian Phase II and human xenograft activities. Literature references are shown in superscript font. *B*, scatter plots and correlation analysis for the same tumor clinical *versus* preclinical activity relationship in ovarian cancer based on the data in Fig. 5A. Analysis was done for (*a*) when all of the xenografts were included in preclinical activity calculations ("All") and (*b*) when only the six xenografts in the gemcitabine panel were used for preclinical activity calculations, where possible ("Gem. Panel"). Stars indicate statistical significance at the 5% level.

might be predictive in the case of typical cytotoxic cancer agents but might fail to provide reliable information for at least some of the noncytotoxic cancer drugs. Additional studies are needed to explore this observation.

The fact that drug potency (mean $Log_{10}GI_{50}$), a pharmacological measure, was found to be predictive of Phase II performance was somewhat surprising but has been noted previously: a recent study by Johnson *et al.* (18) demonstrated a highly significant correlation between potency in the NCI human tumor cell line screen and activity in the hollow fiber assay. Pharmacological considerations (pharmacological differences between the species) might provide a possible explanation why some anticancer agents appear effective in *in vivo* mouse models but fail to show efficacy in Phase II trials. Experience with some agents (57) has shown that the maximum-tolerated dose in mouse can be higher than in humans, presumably because of an intrinsic ability of mouse cells to tolerate higher drug doses and/or more efficient elimination in the mouse.

In contrast to the *in vitro* cell line, our results suggest that the murine allograft model, as used in this analysis, is not predictive of clinical Phase II performance. This is in agreement with the conclusions from a large body of information originating from the NCI screening programs in use from 1975 to 1990 (5–8, 10–12).

The human xenograft model showed good tumor-specific predictive value for NSCLC and ovarian cancers when panels of xenografts were used. However, it failed to adequately predict clinical performance both in the disease and compound-oriented settings for breast and colon tumors. The results with breast cancer were in agreement with a recent study (18) but were contradictory to the work reported by Bailey *et al.* (20), Inoue *et al.* (21), and Mattern *et al.* (24). However, given that the latter studies did not use formal statistical methods, our conclusions may be more robust. The results for ovarian cancer were in agreement with studies by Taetle *et al.* (23) and Mattern *et al.* (24) but contradicted the conclusions of the recent NCI United States study by Johnson *et al.* (18). Our results for NSCLC were consistent with the observations from all previous studies that examined same tumor correlations in this cancer type (18, 24).

For NSCLC and ovarian cancer patients, a panel of xenografts was more predictive than single xenografts confirming preliminary observations by Bellet *et al.* (19).

In an effort to identify the properties that may render an ovarian or NSCLC human xenograft panel predictive of Phase II drug performance, common characteristics were sought. There was no similarity in number and only limited overlap in identity of xenografts between same tumor type panels. However, certain patterns in histology/grade content were found. These ob-

servations suggest that the relative histology/grade content rather than the number or identity of xenografts within a panel may be the important determinant of clinical predictivity. To our knowledge, no other study has attempted to identify ovarian or NSCLC human xenograft panel features that might lead to accurate predictions of a drug's Phase II performance.

This is the only study that has examined the clinical predictive value of three preclinical cancer models together and thus allows for direct comparisons between them. The results suggest that the human xenograft model is more predictive than its murine allograft counterpart and that the *in vitro* cell line model is of, at least, equivalent usefulness to the human xenograft model.

The NCI work with cancer drug screening programs from 1955 to 1990 (Refs. 5–8, 10–12; leukemia-based preclinical, compound-oriented screens preferentially yielding compounds active against hematological malignancies) in combination with our work and recent conclusions by Johnson *et al.* (Ref. 18; statistically significant results under the compound-oriented approach for some solid tumors) suggest that the compound-oriented strategy may be successful when used only within solid tumors or only within hematological malignancies but not when the two disease groups are considered together.

In general, our results suggest that the *in vitro* human tumor cell line and the human xenograft models might have good clinical predictive value in some solid tumors (such as ovary and NSCLC) under both the disease and compound-oriented strategies, as long as an appropriate panel of tumors is used in preclinical testing.

In conclusion, given the results in this study and those of others (6, 7, 10–12), continued use of the murine allograft model in drug development may not be justified. The work presented here argues for emphasis to be placed on *in vitro* cell lines (in the context of the NCI Human Tumor Cell Line Screen) and appropriate panels of the human xenograft model.

Recent years have seen an explosion in the molecular understanding of cancer, which has led to the development of not only more effective cytotoxic cancer drugs but of potentially cytostatic or antimetastatic agents as well. The future preclinical and clinical development of traditional cytotoxic compounds will likely follow similar procedures with those practiced today, and in that sense, the present findings could contribute to the more efficient discovery of such agents. However, the existing cancer models and parameters of activity in both the preclinical and clinical settings may have to be redesigned to fit the mode of action of the novel cytostatic, antimetastatic, antiangiogenesis, or immune response-modulating agents (58). In the preclinical cancer model front, the case is being made for the use of the orthotopic mouse xenograft and transgenic models (59–61) because those are thought to more accurately simulate human disease, especially in terms of growth characteristics and metastatic behavior. New end points of preclinical activity are contemplated such as the demonstration that a new molecule truly hits the intended molecular target (58). In Phase II clinical trials, there is a growing effort toward validating new surrogate endpoints of drug efficacy (58). The next decade will probably answer many of the questions regarding the effectiveness of these novel agents and will likely define a new role for tradi-

tional cytotoxic therapies, but it will also bring new challenges in terms of preclinical predictors of activity.

REFERENCES

1. Curt, G. A. The use of animal models in cancer drug discovery and development. Stem Cells, *12:* 23–29, 1994.

2. Goldin, A., Serpick, A. A., and Mantel, N. Experimental screening procedures and clinical predictability value. Cancer Chemother. Rep., *50:* 173–218, 1966.

3. Venditti, J. M. Drug evaluation branch program: report to screening contractors. Cancer Chemother. Rep., *5:* 1–4, 1975.

4. Venditti, J. M. Preclinical drug development: rationale and methods. Semin. Oncol., *8:* 349–353, 1981.

5. Venditti, J. M., Wesley, R. A., and Plowman, J. Current NCI preclinical antitumor screening *in vivo*: results of tumor panel screening, 1976–1982 and future directions. Adv. Pharmacol. Chemother., *20:* 1–19, 1984.

6. Venditti, J. M. The National Cancer Institute antitumor drug discovery program, current and future perspectives: a commentary. Cancer Treat. Rep., *67:* 767–772, 1983.

7. Staquet, M. J., Byar, D. P., Green, S. B., and Rozencweig, M. Clinical predictivity of transplantable tumor systems in the selection of new drugs for solid tumors: rationale for a three-stage strategy. Cancer Treat. Rep., *67:* 753–765, 1983.

8. Goldin, A., Venditti, J. M., MacDonald, J. S., Muggia, F. M., Henney, J. E., and De Vita V. T. Current results of the screening program at the division of cancer treatment, National Cancer Institute. Eur. J. Cancer, *17:* 129–142, 1981.

9. Marsoni, S., and Wittes, R. Clinical development of anticancer agents: A National Cancer Institute perspective. Cancer Treat. Rep., *68:* 77–85, 1983.

10. De Vita, V. T., and Schein, P. S. The use of drugs in combination for the treatment of cancer N. Engl. J. Med., *288:* 998–1006, 1973.

11. Zubrod, C. G. Chemical control of cancer. Proc. Natl. Acad. Sci. USA, *69:* 1042–1047, 1972.

12. Marsoni, S., Hoth, D., Simon, R., Leyland-Jones, B., De Rosa, M., and Wittes, R. E. Clinical drug development: an analysis of Phase II trials, 1970–1985. Cancer Treat. Rep., *71:* 71–80, 1987.

13. Shoemaker, R. H., Monks, A., Alley, M. C., Scudiero, D. A., Fine, D. L., McLemore, T. L., Abbott, B. J., Paull, K. D., Mayo, J. G., and Boyd, M. R. Development of human tumor cell line panels for use in disease-oriented drug screening. Prog. Clin. Biol. Res., *276:* 265–286, 1988.

14. Skehan, P., Storeng, R., Scudiero, D., Monks A., McMahon, J., Vistica, D., Warren, J. T., Bokesch, H., Kenney, S., and Boyd, M. R. New colorimetric cytotoxicity assay for anticancer drug screening. J. Natl. Cancer Inst. (Bethesda), *82:* 1107–1112, 1990.

15. Rubinstein, L. V., Shoemaker, R. H., Paull, K. D., Simon, R. M., Tosini, S., Skehan, P., Scudiero, D., Monks, A., and Boyd, M. R. Comparison of *in vitro* anticancer drug screening data generated with a tetrazolium assay *versus* a protein assay against a diverse panel of human tumor cell lines. J. Natl. Cancer Inst. (Bethesda), *82:* 1113–1118, 1990.

16. Alley, M. C., Scudiero, D., Monks, A., Hursey, M. L., Czerwinski, M. J., Fine, D. L., Abbott, B. J., Mayo, J. G., Shoemaker, R. H., and Boyd, M. R. Feasibility of drug screening with panels of human tumor Cell lines using a microculture tetrazolium assay. Cancer Res., *48:* 589–601, 1988.

17. Monks, A., Scudiero, D., Skehan, P., Shoemaker, R. H., Paull, K. D., Vistica, D., Hose, C., Langley, J., Cronise, P., Vaigro-Wolff, A., Gray-Goodrich, M., Campbell, H., Mayo, J., and Boyd, M. R. Feasibility of a high-flux anticancer drug screen using a diverse panel of cultured human tumor cell lines. J. Natl. Cancer Inst. (Bethesda), *83:* 757–766, 1990.

18. Johnson, J. I., Decker, S., Zaharevitz, D., Rubinstein, L. V., Venditti, J. M., Schepartz, S., Kalyandrug, S., Christin, M., Arbuck, S., Hollingshead, M., and Sausville, E. A. Relationships between drug

activity in NCI preclinical *in vitro* and *in vivo* models and early clinical trials. Br. J. Cancer, *84:* 1424–1431, 2001.

19. Bellet, R. E., Danna, V., Mastrangelo, M. J., and Berd, D. Evaluation of a "nude" mouse-human tumor panel as a predictive secondary screen for cancer chemotherapy agents. J. Natl. Cancer Inst. (Bethesda), *63:* 1185–1187, 1979.

20. Bailey, M. J., Gazet, J-C., Smith, I. E., and Steel, G. G. Chemotherapy of human breast-carcinoma xenografts. Br. J. Cancer, *42:* 530–536, 1980.

21. Inoue, K., Fujimote, S., and Ogawa, M. Antitumor efficacy of seventeen anticancer drugs in human breast cancer xenograft (MX-1) transplanted in nude mice. Cancer Chemother. Pharmacol., *10:* 182–186, 1983.

22. Steel, G. G., Courtenay, V. D., and Peckhan M. J. The response to chemotherapy of a variety of human tumor xenografts. Br. J. Cancer, *47:* 1–13, 1983.

23. Taetle, R., Rosen, F., Abramson, I., Venditti, J., and Howell, S. Use of nude mouse xenografts as preclinical drug screens: *in vivo* activity of established chemotherapeutic agents against melanoma and ovarian carcinoma xenografts. Cancer Treat. Rep., *71:* 297–304, 1987.

24. Mattern, J., Bak, M., Hahn, E. W., and Volm, M. Human tumor xenografts as models for drug testing. Cancer Metastasis Rev., *7:* 263–284, 1988.

25. Boven, E., Winograd, B., Fodstad, O., Lobbezzoo, M. W., and Pinedo, H. M. Preclinical Phase II studies in human tumor lines: a European multicenter study. Eur. J. Cancer, *24:* 567–573, 1988.

26. Boven, E., Winograd, B., Berger, D. P., Dumant, M. P., Braakhuis, B. J. M., Fodstad, O., Langdon, S., and Fiebig, H. H. Phase II preclinical drug screening in human tumor xenografts: a first European multicenter collaborative study. Cancer Res., *52:* 5940–5947, 1992.

27. Langdon, S., Hendriks, H. R., Braakhuis, B. J. M., Pratesi, G., Berger, D. P., Fodstad, O., Fiebig, H., and Boven, E. Preclinical Phase II studies in human tumor xenografts: a European multicenter follow-up study. Ann. Oncol., *5:* 415–422, 1994.

28. Paull, K. D., Shoemaker, R. H., Hodes, L., Monks, A., Scudiero, D., Rubinstein, L. V., Plowman, J., and Boyd, M. R. Display and analysis of patterns of differential activities of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm. J. Natl. Cancer Inst. (Bethesda), *81:* 1088–1092, 1989.

29. Rosner, B. Non-parametric Methods. *In:* Fundamentals of Biostatistics, 4th ed., p. 575. Belmont, CA: Duxbury Press, Wadsworth Publishing Company, 1994.

30. Kolfscoten, G. M., Pinedo, H. M., Scheffer, P. G., Schluper, H. M. M., Erkelens, C. A. M., and Boven, E. Development of a panel of 15 human ovarian cancer xenografts for drug screening and determination of the role of the glutathione detoxification system. Gynecol. Oncol., *76:* 362–368, 2000.

31. Louie, A. C., and Issell, B. F. Amsacrine (AMSA): a clinical review. J. Clin. Oncol., *3:* 562–592, 1985.

32. Minna, J. D., Pass, H., Glatsein, E., and Inde, D. Cancer of the lung. *In:* V. T. DeVita, S. Hellman, and S. A. Rosenberg (eds.), Cancer Principles and Practice of Oncology, pp. 660 (Table 22–48). Philadelphia: J. B. Lippincott, 1989.

33. Gordon, A. N., Granai, C. O., Rose, P. G., Hainsworth, J., Lopez, A., Weissman, C., Rosales, R., and Sharpington, T. Phase II study of liposomal doxorubicin in platinum- and paclitaxel-refractory epithelial ovarian cancer. Clin. Oncol., *18:* 3093–3100, 2000.

34. Israel, V. P., Garcia, A. A., Roman, L., Muderspach, L., Burnett, A., Jeffers, S., and Muggia, F. M. Phase II study of liposomal doxorubicin in advanced gynecologic cancers. Gynecol. Oncol., *78:* 143–147, 2000.

35. Muggia, F. M., Hainsworth, J. D., Jeffers, S., Miller, P., Groshen, S., Tan, M., Roman, L., Uziely, B., Muderspach, L., Garcia, A., Burnett, A., Greco, F. A., Morrow, C. P., Paradiso, L. J., and Liang, L. J. Phase II study of liposomal doxorubicin in refractory ovarian cancer: antitumor activity and toxicity modification by liposomal encapsulation. J. Clin. Oncol., *15:* 987–993, 1997.

36. Blum, R. H., and Carter, S. K. Adriamycin: a new anticancer drug with significant clinical activity. Ann. Intern. Med., *80:* 249–259, 1974.

37. Thigpen, J. T., Lagasse, L., Homesley, H., and Blessing, J. A. *Cis*-platinum in the treatment of advanced or recurrent adenocarcinoma of the ovary. A Phase II study of the Gynecology Oncology. Group. Am. J. Clin. Oncol., *6:* 431–435, 1983.

38. Wiltshaw, E., and Kroner, T. Phase II study of *cis*-dichlorodiammineplatinum(II) (NSC-119875) in advanced adenocarcinoma of the ovary. Cancer Treat. Rep., *60:* 55–60, 1976.

39. Bonomi, P. D., Mladineo, J., Morrin, B., Wilbanks, G., Jr., and Slayton, R. E. Phase II trial of hexamethylmelamine in ovarian carcinoma resistant to alkylating agents. Cancer Treat. Rep., *63:* 137–138, 1979.

40. Johnson, B. L., Fisher, R. I., Bender, R. A., DeVita, V. T., Jr., Chabner, B. A., and Young, R. C. Hexamethylmelamine in alkylating agent-resistant ovarian carcinoma. Cancer (Phila.), *42:* 2157–2161, 1978.

41. Omura, G. A., Blessing, J. A., Morrow, C. P., Buchsbaum, H. J., and Homesley, H. D. Follow-up on a randomized trial of melphalan (M) vs. melphalan plus hexamethylamine (M+H) *versus* Adriamycin plus cyclophosphamide (A+C) in advanced ovarian carcinoma. Proc. Am. Assoc. Cancer Res., *22:* 470, 1981.

42. Bolis, G., D'Incalci, M., Belloni, C., and Mangioni, C. Hexamethylmelamine in ovarian cancer resistant to cyclophosphamide and Adriamycin. Cancer Treat. Rep., *63:* 1375–1377, 1979.

43. Manetta, A., MacNeill, C., Lyter, J. A., Scheffler, B., Podczaski, E. S., Larson, J. E., and Schein, P. Hexamethylmelamine as a single second-line agent in ovarian cancer. Gynecol. Oncol., *36:* 93–96, 1990.

44. Rosen, G. F., Lurain, J. R., and Newton, M. Hexamethylmelamine in ovarian cancer after failure of cisplatin-based multiple-agent chemotherapy. Gynecol. Oncol., *27:* 173–179, 1987.

45. Parker, L. M., Griffiths, C. T., Yankee, R. A., Knapp, R. C., and Canellos, G. P. High-dose methotrexate with leucovorin rescue in ovarian cancer: a Phase II study. Cancer Treat. Rep., *63:* 275–279, 1979.

46. Barlow, J. J., and Piver, M. S. Methotrexate (NSC-740) with citrovorum factor (NSC-3590) rescue, alone and in combination with cyclophosphamide (NSC-26271), in ovarian cancer. Cancer Treat. Rep., *60:* 527–533, 1976.

47. Morgan, R. J., Jr., Speyer, J., Doroshow, J. H., Margolin, K., Raschko, J., Sorich, J., Akman, S., Leong, L., Somlo, G., and Vasilev, S. Modulation of 5-fluorouracil with high-dose leucovorin calcium: activity in ovarian cancer and correlation with CA-125 levels. Gynecol. Oncol., *58:* 79–85, 1995.

48. Markman, M., Reichman, B., Hakes, T., Hoskins, W., Rubin, S., Jones, W., and Lewis, J. L., Jr. Intraperitoneal chemotherapy as treatment for ovarian carcinoma and gastrointestinal malignancies: the Memorial Sloan-Kettering Cancer Center experience). Acta Med. Austriaca, *16:* 65–67, 1989.

49. Prefontaine, M., Donovan, J. T., Powell, J. L., and Buley, L. Treatment of refractory ovarian cancer with 5-fluorouracil and leucovorin. Gynecol. Oncol., *61:* 249–252, 1996.

50. Kamphuis, J. T., Huider, M. C., Ras, G. J., Verhagen, C. A., Kateman, I., Vreeswijk, J. H., and Burghouts, J. T. High-dose 5-fluorouracil and leucovorin as second-line chemotherapy in patients with platinum-resistant epithelial ovarian cancer. Cancer Chemother. Pharmacol., *37:* 190–192, 1995.

51. Burnett, A. F., Barter, J. F., Potkul, R. K., Jarvis, T., and Barnes, W. A. Ineffectiveness of continuous 5-fluorouracil as salvage therapy for ovarian cancer. Am. J. Clin. Oncol., *17:* 490–493, 1994.

52. Reed, E., Jacob, J., Ozols, R. F., Young, R. C., and Allegra, C. 5-Fluorouracil (5-FU) and leucovorin in platinum-refractory advanced stage ovarian carcinoma. Gynecol. Oncol., *46:* 326–329, 1992.

53. Ozols, R. F., Speyer, J. L., Jenkins, J., and Myers, C. E. Phase II trial of 5-FU administered Ip to patients with refractory ovarian cancer. Cancer Treat. Rep., *68:* 1229–1232, 1984.

54. Long, H. J., III, Nelimark, R. A., Su, J. Q., Garneau, S. C., Levitt, R., Goldberg, R. M., Poon, M. A., and Kugler, J. W. Phase II evaluation of 5-fluorouracil and low-dose leucovorin in cisplatin-refractory advanced ovarian carcinoma. Gynecol. Oncol., *54:* 180–183, 1994.

55. Look, K. Y., Muss, H. B., Blessing, J. A., and Morris, M. A. Phase II trial of 5-fluorouracil and high-dose leucovorin in recurrent epithelial ovarian carcinoma. A Gynecology Oncology Group Study. Am. J. Clin. Oncol., *18:* 19–22, 1995.

56. De Graeff, A., van Hoef, M. E., Tjia, P., Heintz, A. P., and Neijt, J. P. Continuous infusion of 5-fluorouracil in ovarian cancer patients refractory to cisplatin and carboplatin. Ann. Oncol., *2:* 691–692, 1991.

57. Volpe, D. A., Tomaszewski, J. E., Parchment, R. E., Garg, A., Flora, K. P., Murphy, M. J., and Grieshaber, C. K. Myelotoxic effects of the bifunctional alkylating agent bizelesin in human, canine and murine myeloid progenitor cells. Cancer Chemother. Pharmacol., *39:* 143, 1996.

58. Gelmon, K. A., Eisenhauer, E. A., Harris, A. L., Ratain, M. J., and Workman, P. Anticancer agents targeting signaling molecules and cancer cell environment: challenges for drug development? J. Natl. Cancer Inst. (Bethesda), *91:* 1281–1287, 1999.

59. Killion, J. J., Radinsky, R., and Fidler, I. J. Orthotopic models are necessary to predict therapy of transplantable tumors in mice. Cancer Metastasis Rev., *17:* 279–284, 1999.

60. Kerbel, R. S. What is the optimal rodent model for anti-tumor drug testing? Cancer Metastasis Rev., *17:* 301–304, 1999.

61. Rosenberg, M. P., and Bortner, D. Why transgenic and knockout animal models should be used (for drug efficacy studies in cancer). Cancer Metastasis Rev., *17:* 295–299, 1999.